DOCUMENT RESUME

ED 136 425 EA 009 349

AUTHOR Owens, Thomas R.; Hiscox, Michael D.
TITLE Alternative Models for Adversary Evaluation: Variations on a Theme.
PUB DATE Apr 77
NOTE 29p.; Paper presented at the Annual Meeting of the American Educational Research Association (New York, N.Y., April 4-8 1977); Pages 23-24 may not reproduce clearly due to small print size

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS Case Studies (Education); Elementary Secondary Education; *Evaluation Methods; Higher Education; Models
IDENTIFIERS *Adversary Evaluation

ABSTRACT

Growing dissatisfaction with conventional evaluation approaches to dealing with complex social phenomena has led evaluators to search out alternative evaluation models. Exclusive reliance on the experimental model and the logic of statistical inference has been seriously attacked on philosophic, technical, and pragmatic grounds. This paper deals with existing problems in traditional evaluation and presents a rationale for exploring alternative approaches, along with a brief case study description of six applications of adversary evaluation using the judicial and debate forms. Some common features noted in the six adversary evaluation case studies were 1) better communication between evaluators and decision makers, 2) greater attention to the formulation of key evaluation issues, and 3) increased concern for meta-evaluation. The authors conclude with seven unresolved issues in adversary evaluation. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED136425

ALTERNATIVE MODELS FOR ADVERSARY EVALUATION: VARIATIONS ON A THEME

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Thomas R. Owens
and
Michael D. Hiscox

Northwest Regional Educational Laboratory
710 S.W. Second Avenue
Portland, Oregon 97204

Paper Presented at the American Educational Research Association
Annual Meeting in New York City, New York, April 1977

EA 009 349

ALTERNATIVE MODELS FOR ADVERSARY EVALUATION: VARIATIONS ON A THEME

Thomas R. Owens and Michael D. Hiscox
Northwest Regional Educational Laboratory, Portland, Oregon

Growing dissatisfaction with conventional evaluation approaches to dealing with complex social phenomena has led evaluators to search out alternative evaluation models. Exclusive reliance on the experimental model and the logic of statistical inference has been seriously attacked on philosophic, technical, and pragmatic grounds. This paper deals with existing problems in traditional evaluation and presents a rationale for exploring alternative approaches, along with a brief history and description of some adversary models of evaluation, some common features of alternative approaches, and some unresolved issues in adversary evaluation.

Why all the Fuss?

In struggling with the problems of providing useful information to decision makers and making reasonable judgments about the merits of complex educational and social programs, evaluators have become increasingly aware of the limitations of experimental evaluation designs. As Levine (1974) has said,

> The experiment cannot deal with historical contexts, and it requires the reduction of whole human events to be contrived dimensions that can be quantified. Moreover, the transactional nature of human interaction precludes the assumptions that experiments make about humans as interchangeable, reactive creatures who each respond to the stimulus and the meaning for it the experimenter had in mind.

> In the search for methodological purity, social scientists have often lost sight of the substantive problems the methods were meant to solve. (p. 674)

House (1976, p. 1) expresses dissatisfaction with reliance on the traditional "student-gain-by-testing" approach on philosophic grounds. He believes that such an approach is based on utilitarian ethics which aim to "achieve the greatest net balance of satisfaction as summed over all individuals" using a common index of satisfaction (generally, standardized test scores). The utilitarian ethic ignores the need to consider multiple outcomes and measures, emphasizes the magnitude of test score gains without consideration for processes used to achieve them, ignores the pluralistic values of diverse groups interested in the evaluation, and often favors the growth of the higher social classes at the expense of the lower.

On technical grounds, "the logic of experimentation and of statistical inference demands conditions that are difficult, if not impossible, to meet in human studies." (Levine, 1974, p. 663) Problems include the use of comparison group data, the presence of extraneous variables, experimenter effects, the limitations of existing measurement instruments, and the generalization of findings to new settings.

From a pragmatic point of view, many evaluation studies simply end up on someone's shelf because they are unresponsive to the needs of people involved in or affected by a program being evaluated. Wolf (1975) states that

> This occurs because most of the current methods are rooted in behavioral and social science research and rely on quantification and technical analysis. Great collections of numbers, such as those found in children's

cumulative files and school or program evaluation studies, tend to blur and obscure rather than sharpen and illuminate the education process. In seeking objectivity, the decision maker using these methodologies may exclude a factor that ought to be of fundamental concern: human judgment. (p. 185)

What's Involved in Adversary Evaluation?

Dissatisfaction with traditional evaluation processes in education has led to the exploration of new approaches borrowed and adapted from such fields as accounting, anthropology, economics, law and politics. This paper discusses recent developments in the use of adversary models in educational evaluation. These models borrow freely from legal procedures used in jury trials and administrative hearings and from principles of public debate. Specific applications of adversary evaluations are described in the next section of this paper. Before considering alternative approaches, however, it may be helpful to review some common assumptions supporting these approaches.

The following seven assumptions appear directly related to the adversary models of evaluation described in this paper.

1. The fundamental evaluation instrument is the human intelligence deciphering complex data.
2. The evaluator is not a purely "rationale and impartial spectator," but is subject to certain biases that require controls.
3. Social and educational phenomena are multidimensional; therefore, an effective evaluation must be responsive to these various dimensions. Qualitative as well as quantitative data are needed.
4. Evaluation occurs in a pluralistic society in which differing value perspectives must be addressed.

5. Decision makers are interested in considering alternative interpretations of evaluation data.

6. An adversary approach to evaluation can uncover some significant insights into phenomena that are important for decision making but would likely be overlooked in a traditional study by an "objective" evaluator.

7. Important decisions regarding large scale programs are seldom made by a single individual. Therefore, an effective evaluation needs to incorporate a wide range of input from various people and to communicate the findings and interpretations to a broad audience.

What Alternative Approaches to Adversary Evaluation Have Been Applied?

This section provides case study summaries of six adversary evaluations. In working with the model, we have become aware of several uses for adversary methods. In addition, we surveyed eleven other evaluators throughout the country for information on additional case studies. It is probable, nevertheless, that we have overlooked other good examples of adversary evaluations.

1. Man: A Course of Study

The earliest reported application of an adversary approach to evaluation occurred in the form of an experimental hearing held at the Hawaii Curriculum Center at the University of Hawaii in February 1970. The purpose of the experiment was to examine the usefulness of a modified judicial model--the administrative adversary hearing--as an aid to curriculum evaluation. The focus was on

> exploring the technique rather than...the substantive concern, so a hypothetical issue was used. The hearing was to decide whether the curriculum Man: A Course

> of Study, developed by the Educational Development Center in Massachusetts, should be adopted for the public schools in Hawaii. (Owens, 1973, p. 300)

Two educators with experience in teaching social science served as advocate and adversary regarding potential curriculum adoption. In preparing their cases, both selected and interviewed witnesses and studied existing documents related to this curriculum and the Department of Education's social studies guide.

> The hearings officer organized the pretrial conferences with the defense and prosecution, established flexible rules of operation, required that a written outline of the arguments against the adoption of the curriculum be given to the defense and to himself before the hearing, conducted the hearing, ruled on the admissibility of evidence and the propriety and relevance of questions if objectives were raised by either party, limited the hearing to two hours, and specified which points were in contention at any time during the hearing. (p. 301)

A representative jury consisted of a public school sixth grade social studies teacher, the State Department of Education's program specialist in social studies and a university evaluation officer.

Approximately 40 members of the university faculty and the State Department of Education accepted an open invitation to observe and comment on the hearing. Their consensus was that the two-hour hearing served as an excellent way of clarifying many facets of a new curriculum. Nevertheless, the majority felt that two hours was insufficient time in which to reach a verdict.

2. Experience-Based Career Education

In the summer of 1974, a prototype adversary hearing was conducted by the Career Education Program of the Northwest Regional Educational Laboratory (NWREL) in Portland, Oregon. At issue was the question of whether Experience-Based Career Education (EBCE) should be adopted by school districts in the Pacific Northwest. EBCE is a new approach to secondary education, developed by the National Institute of Education and four regional educational laboratories to help bridge the gap between the classroom and the community. EBCE students learn basic skills, life skills and career development skills through direct experience with adults in the community.

As development of the EBCE program stabilized and the program moved into a replication and dissemination phase, the need grew for an evaluation mechanism that would help potential adopters recognize the strengths and weaknesses of the program. It was felt that a traditional evaluation report would not present enough balanced information on those issues important to potential adopters. Therefore, the decision was made to use an adversary approach.

A videotape of this adversary hearing was to be made to aid a school district in deciding whether to adopt EBCE. But the hearing was intended as only a starting point in refining the application of an adversary approach to educational evaluation. Many procedures of a complete hearing were not followed precisely because of limited resources or lack of understanding regarding how to apply them. For example, the charges were not clearly specified, rules of evidence were not fully developed, only four witnesses were used and the processes of jury participation and deliveration were omitted. Nevertheless, certain features of the judicial process (such as preliminary rules of evidence, pretrial discovery, interviewing of witnesses and examination of evidence, use of a hearings officer, testimony and cross-examination of witnesses, and initial

and concluding statements of the advocate and adversary) were employed and adopted to program evaluation.

As a result of a planning conference, a set of guidelines was drafted for use by the project coordinators in producing the hearing videotape.

Over the next few weeks, these guidelines evolved into the actual specifications for production of the trial. The most important of these specifications are noted here.

1. After introductory presentations about EBCE, several potential adopter groups of educators were asked to list special strengths and weaknesses of the EBCE program. The most frequent responses were then put on a questionnaire that was completed by a different group of potential adopters. On this questionnaire specific issues concerning EBCE were stated, and respondents were instructed to identify, on a five-point rating scale, the extent to which they felt each issue was a strength or a weakness of EBCE. Based on this feedback, the evaluators identified the five principal positive and the five principal negative points of contention for the trial. For each of these ten points of contention, project personnel identified witnesses, possible lines of testimony and specific supporting or rejecting evidence for use by the advocate and adversary.

2. The advocate and adversary, both educators, were selected on the basis of interest in the adversary method, independence from the EBCE project, ability to communicate effectively, and necessary skill to quickly and accurately evaluate the essential aspects of an educational program. Dr. Wolf of Indiana University and Dr. Terry Denny of the University of Illinois were chosen.

3. A local law school professor was selected as hearings officer. His input to the overall production was valuable but minor because of his limited role (mainly that of presiding over the courtroom trial). However, he did provide a very helpful critique of the proposed legal techniques.

4. A great deal of effort over a four-month period went into selecting the most important trial issues, verifying the selection with school administrators, and formulating arguments and collecting evidence. The total information accumulated was probably sufficient for a trial of several days. However, this information was used sparingly because of our decision to produce only a one-hour videotape of the prototype hearing.

5. Coordinators developed guidelines which covered rules of evidence, cross-examination techniques, and so forth. Though their work provided an interesting opportunity to explore the intricacies of the legal model, the prototype hearing produced was not sufficiently complex for these guidelines to have any noticeable impact.

6. As the trial date approached, it became obvious that a one-hour trial would not allow time for sufficient discussion of issues, nor would it permit more than a superficial deliberation of the adoption question. Therefore, plans for a jury deliberation were dropped. It was also decided that some prior knowledge, gained from brochures, site visits or project orientation sessions, would be required of potential EBCE adopters before watching the videotape would be beneficial.

The hearing was conducted under the following conditions:

1. Four witnesses (the EBCE project director, learning manager, cooperating employer and former student) would be used. The advocate and adversary had selected and jointly interviewed each witness prior to the hearing. It was felt that using only four witnesses would add to the continuity of the hearing, and serve to illustrate various segments of a longer hearing.
2. A proposed schedule specified the time available for each segment of the hearing in order to produce one hour of videotape. If the presenters had not finished with a witness (and it was not expected that they would) in the time allotted, the testimony would be "faded out" by the technical director.
3. No jury was present, although some people were asked to represent a jury for camera effect.
4. The four witnesses were to represent those favorable to EBCE adoption. Each would be cross-examined by the adversary, but because of time limitations, no hostile witnesses were scheduled. The hearing would then jump to the advocate and adversary's closing arguments, instructions from the hearings officer to the jury and a narrator's final comments.

Following the pretrial conference and preliminary deposition of witnesses by the advocate and adversary, a hearing was held in Portland, Oregon, in July 1974. Because of various staging and technical production problems, it took three and a half hours to produce a one-hour edited tape.

During the actual one-hour hearing, the advocate for adoption stressed the program's positive effects on students, and the increased implementation knowledge that would allow NWREL staff to provide adequate technical support

to districts wishing to adopt the program. The adversary proposed that the program was working well in Tigard (the demonstration site) primarily because of the skills and enthusiasm of program staff, rather than as a result of any factors inherent in program design. He cautioned the jury that fine staff do not come with the program, and that it would be difficult to find similarly qualified staff in most local school districts.

3. Independent Plans of Study

In the fall of 1974, Terry Denny and his graduate class conducted an adversary hearing at the University of Illinois in an experimental undergraduate liberal arts program known as Independent Plans of Study (IPS). Stenzel (1976a) stated "IPS was near the end of its term as an experimental project and would have to show proof of its success if it were to continue. The adversary hearing would serve as a way of giving a forecast of the best and worst that might be found." (p. 8) Two issues were identified for the hearing: "Does IPS produce satisfied students who meet departmental standards?" and "Putting economic considerations aside, does the IPS program merit continued existence?" (pp. 9-10)

Data were collected cooperatively by graduate students on advocate and adversary teams, who conducted a records search, distributed questionnaires to current and past students, and interviewed relevant parties. After data were reviewed by both teams, it was jointly agreed that IPS produced satisfied students who meet departmental standards. Since there was no contention of this issue, the question of the IPS program's continued existence was selected as the issue for the hearing.

The jury trial metaphor was abandoned because of perceived difficulty in obtaining witnesses and because much of the evidence would have been introduced by the teams themselves. An appellate approach was substituted, using

a debate style format in which arguments and rebuttals were presented by both teams and the panel of judges was allowed to ask questions.

Norman Stenzel, an educational specialist from the Illinois Office of Education, was employed to conduct a meta-evaluation (an evaluation of the evaluation) of both the IPS hearing and the Indiana University Division of Teacher Education hearing described in the next section of this paper. The IPS meta-evaluation was based on structured interviews with persons who had participated in the hearing. Stenzel found mixed reactions among those interviewed regarding the significance of issues and the effectiveness of processes used (Stenzel, 1976b).

4. Division of Teacher Education

The first full-scale test of the judicial evaluation model occurred in May 1975. The model was used to assess the improvement of teacher education through programs of the Division of Teacher Education (DTE) at Indiana University.

> Issues were identified upon which the inquiry could focus. Adversary teams built and presented their arguments for and against the program. Thirty-two witnesses testified over a two-day period. Documents were entered into the record. Strategies for direct, cross-, redirect, and recross-examination were developed. Judicial instructions, rules of evidence, jury selection, deliberation procedures, and prehearing discovery were all designed and implemented. The jury panel (13 in all) was comprised of educational experts from around the country, including the program's monitor from the U.S. Office

> of Education and faculty and students from Indiana University. Basing judgment on the evidence presented, the panel rendered judgments on the issues identified and offered prescriptions for modification and reform. (Wolf, 1975, p. 186)

The Division of Teacher Education consists of over 30 alternative teacher preparation projects. According to Wolf, the intent of the evaluation was to make a summative judgment about the DTE, and to provide a basis for future planning and decision making.

> The judicial evaluation model was selected partly because it would involve a wide spectrum of people. Judgment was not to rest solely on the program's original objectives. The examination would focus on relevant issues. The model offered a format to assist in decision making; it relied on the broad information base that existed in the DTE. And, most significantly, it promised to present a comprehensive view of the program, to illuminate strengths as well as weaknesses, and to accomplish all of this in a public fashion. (p. 186)

The judicial evaluation of the DTE took six months and involved four stages: issue generation, issue selection, preparation of arguments, and the hearing itself. Through interviews with faculty, students and individuals from the funding agency, over 30 potential issues were identified. This number was later reduced through a survey instrument in which faculty, students and administrators established priority among issues. In stage three, specific points of contention were developed concerning three primary

issues: (1) the effectiveness of the organizational structure under which DTE programs operated, (2) the difference between DTE programs and what had existed previously, and (3) the extent of the conceptual bases for teacher education programs. (Stenzel, 1976b, p. 12)

In stage four, ground rules were established to cover the hearing and the learning itself. The hearing consisted of three half-day sessions and a half day for jury deliberation. The opposing counsels, Robert Wolf and Roger Farr, were evaluators from the Indiana University College of Education. A hearings officer, familiar with legal proceedings and education, served as judge. The jury included faculty and students from Indiana University, outside education experts, and the federal projects officer. Wolf reported that "the DTE evaluation was judged to be quite useful by the program's administrators. Not only did it provide them with a better view of their program, but it greatly facilitated communication." (p. 187)

A meta-evaluation of the DTE hearing was conducted by Norman Stenzel (1975) through personal interviews and tape recorded responses to questions by various participants and observers of the hearing. Some of Stenzel's conclusions based on this meta-evaluation are listed below.

1. A wide range of evidence can be presented in an adversary hearing.
2. The adversary process provides a means of testing evidence and inferences based on that evidence.
3. The derivation of issues is fundamental to a worthwhile hearing.
4. The wording of the issues and the charge to the jury is critical.

5. The familiarity of both the panel and the counsels with the subject of the evaluation bears heavily on how the proceedings are conducted.

6. A balance in strength and skills between the advocate and adversary is a necessity.

7. Because the evaluation must fit diverse situations, the jury trial approach to adversary evaluation should be considered one of a variety of possible approaches. (pp. 20-25)

5. Graduate Program in Clinical-Community Psychology

On April 9 and 10, 1976, a jury trial was conducted as part of an experimental program evaluation at the State University of New York (SUNY) in Buffalo.

> The program being evaluated was the procedure used in place of an examination by the Graduate Program in Clinical-Community Psychology of the Department of Psychology, SUNY Buffalo, to evaluate candidates for the Ph.D. degree....The innovative aspects of the new candidacy procedure were the two papers in place of an examination, and the deadline and penalty procedure. Students were also allowed to select their own committees and to select their own topics in consultation with faculty committees. (Levine, 1976, p. 4)

An interdisciplinary group--lawyers, psychologists, statisticians, educators, and philosophers--studied the problems in applying the jury trial approach to evaluation and in planning the trial. After preparing a statement of standards and purposes, they issued a formal complaint stating that the program violated those standards. The complaint and the response to the complaint became the issues for the trial.

The opposing counsel (advanced law students) were each assisted by teams of three graduate students who worked as case aides. The two teams shared information but developed separate trial strategies. Evidence was required to be relevant. Examination and cross-examination of witnesses was part of the trial. The jury, consisting of faculty and graduate students, was allowed to accress a limited number of questions to witnesses. They were asked to reach a consensus on the validity of each of four questions addressed in the complaint, and to make recommendations for change. The jurors were nearly unanimous in their votes, and the program was found "not guilty" on most charges.

6. The Hawaii 3 on 2 Evaluation

In January 1977, adversary techniques were applied in a new way in evaluating the Hawaii 3 on 2 program. Thus, a more detailed description is given of this case study. The 3 on 2 program is a nine-year old educational innovation in which a team of three professional teachers works with 50 to 70 primary students. The class generally includes children from two consecutive grades, K-1, 1-2, or 2-3. "The goal of the program is to maximize the intellectual, social, emotional and physical growth of students by providing for greater individualization through the team approach to teaching." (NWREL, 1977a, p. 2)

There are 487 3 on 2 classrooms in Hawaii; over half of the K-3 students in the state are in such classrooms. The remainder attend self-contained classrooms where one teacher provides instruction. The 3 on 2 program costs approximately ten million dollars a year for the extra teachers required. For this reason, educators, parents and other taxpayers expressed a desire to measure the program's effectiveness.

Three previous evaluations had been labeled "inconclusive" largely because they focused on student test data and found no significant differences in performance of students who had been in 3 on 2 and self-contained classes. Therefore, in 1976, the Hawaii Department of Education contracted with the Northwest Regional Educational Laboratory to provide a more comprehensive evaluation of 3 on 2.

> Several factors favored the use of an advocate-adversary approach for the 3 on 2 evaluation. First, as noted earlier, the 3 on 2 program is highly visible. It has many supporters--as well as many opponents--throughout the state. Many people feel strongly--pro or con--about the prospect of change in the 3 on 2 program. The advocate-adversary approach provides an optional format for eliciting and reporting data from both viewpoints. This approach also assures decision makers that all sides of the issue have been studied and all relevant data have been presented. (NWREL, 1977b, pp. 10-11)

The evaluation began in March 1976. Two teams of four experienced evaluators each were selected. Each team included two NWREL evaluators and two outside evaluators from other parts of the United States and Canada. Two additional NWREL evaluators served as project co-directors, neutral arbitrators who coordinated the two teams, provided for overall management of the study, and assigned members to the two teams. Teams were assigned to either the advocate or adversary position by the toss of a coin. It was agreed in advance that teams would collect and review a common set of data. The teams agreed on types of

information to be collected and the design to be used. They analyzed data jointly, but interpretation of results was handled in an adversarial manner.

Methods of data collection included content analysis of documents; use of standardized and criterion referenced tests; administration of affective tests to students; questionnaires to parents, teachers, and principals; interviews with hundreds of parents, teachers, school administrators, Hawaii Department of Education staff, members of the Board of Education, state executives, legislators, news reporters and members of the community; "town hall" meetings on four islands to obtain community input; classroom observations; and cost analyses.

An extended site visit was scheduled in May 1976. During that period both teams collected preliminary data about the program and identified evaluation issues important to educators, parents, Board of Education members, legislators and members of the community.

Questionnaires distributed to parents, teachers and principals were designed in an adversary manner. Each team identified arguments they wished to advance and submitted topics and draft items. Several members of each team met to review and argue over each questionnaire item. The intent was to obtain a balance of items reflecting alternate points of view. A similar approach was used in conducting interviews. A member of each team was present during most interviews to insure a balance of questions and to serve as a check on the accuracy of any quotations to be used.

In October, the evaluation teams made another site visit to collect more specific interview data that could support their now refined arguments and to conduct classroom observations in an adversarial manner. The advocate team, having arranged in advance to collect nominations of the best 3 on 2 teams, observed only top-rated 3 on 2 classrooms. The adversary team, on the other

hand, selected a random sample of 3 on 2 classrooms and attempted to use a systematic classroom observation system developed by a national expert in team teaching and individualized instruction. It was clear that the bias of each team dictated its strategy for classroom observations.

To provide a focus the adversary evaluation, one team was charged with arguing that the Hawaii 3 on 2 program should be continued at approximately the same level while the other team was to argue that the program should be severely reduced or eliminated. A debate model, rather than a judicial hearing model, was selected for the presentation. It was felt that a debate format would be more flexible, would allow a broader coverage of issues, and would provide for more efficient use of the limited time available for presentation (about one and one-half hours).

The leader of each team presented his arguments in a closed meeting for Department of Education staff, the Board of Education, the governor and selected legislators. After the two 25-minute presentations, there was a ten-minute rebuttal, followed by questions from the audience. That evening a one-hour presentation was made for television. The televised presentation was carried live throughout the state, and a videotape version shown on the weekend. Following the televised presentations, a ten-day public reaction period was scheduled in which interested citizens could register their reactions and comments to the Department of Education by phone or mail. After the public reaction period, the state superintendent made his recommendations to the Board of Education for their final action.

In addition to the oral debate and television presentation, three volumes were prepared by NWREL: a comprehensive Technical Report describing the evaluation design, data collection and analysis procedures, and results; Team

Interpretations and Recommendations, containing the major arguments for and against 3 on 2; and an Executive Summary providing an overview of the procedures, findings, interpretations and recommendations.

In addition to the televised adversary hearings, the evaluation received widespread newspaper coverage in Hawaii. Editorials were written, legislators made statements, and the evaluation generated a level of public interest that few evaluations could claim. After receiving considerable public input, the Superintendent announced his recommendation that the 3 on 2 program be discontinued. However, in a widely attended and reported five and a half hour meeting, the Board of Education rejected the Superintendent's recommendation, voting five to four to maintain 3 on 2 as an option in Hawaii elementary schools.

Reaction to the evaluation was generally favorable. The number of people applauding the use of the adversary model greatly outweighed those who spoke against it. Those most critical of the model were the evaluators themselves, many of whom thought the polarized structure led either to misleading presentations, an incomplete (in terms of data presented) evaluation, or a compromising of professional responsibilities. However, most audiences seemed delighted with the structure.

Other Approaches

The examples of adversary evaluations cited thus far incorporated a judicial model or a debate model. Other approaches to adversary evaluation have been identified. For example, in what Norman Stenzel (1975) calls a "position paper approach, Robert Stake (1975) and Terry Denny have written an adversary and an advocate statement as part of the evaluation of the Twin City Institute for Talented Youth. Their example of this approach seems limited, however, since the adversary and advocate failed to address the same issues, did not rebut each other's statements, and failed to plan and collect data in advance to support their assertions for the strengths and weaknesses of the program.

Stenzel and others have also proposed a Congressional hearings format in which witnesses could be called to testify on the issues before a panel. Panel members would have a chance to question witnesses before reaching a decision. The authors are not familiar with situations in which a Congressional hearings approach has been employed in educational evaluation.

Some Common Features of Adversary Approaches

A summary of some of the features of the previously described examples of adversary evaluation is provided in the table on the following page under the headings: Purpose, Type of Adversary Model Used, Issue Identification and Selection, Data Collection for Argument Preparation, Presentation (procedures and format), and Decision Making (procedures and criteria).

Spinoff Effects of Adversary Evaluation

Other presenters in this symposium will address the strengths and weaknesses of adversary approaches to evaluation. We would like to share here some observations regarding three spinoff effects of recent adversary evaluations: (1) better communication between evaluators and decision makers, (2) greater attention to formulating key evaluation issues, and (3) increased concern for meta-evaluation.

In each of the six case studies described in this paper, communication between evaluators and decision makers has appeared more effective than in most traditional approaches to evaluation. The need to work together in focusing the evaluation, in collecting data and in reporting the results seems to increase both the opportunity and the need for communication.

Each of the case studies is characterized by concentrated attention on identifying, selecting, and developing evaluation issues. In most cases, a variety of people have been consulted in selecting issues, and particular attention has been given to insuring the relevance of issues to program

CHARACTERISTICS OF SELECTED ADVERSARY EVALUATIONS

Case Study	Purpose	Type of Adversary Model	Issue Identification and Selection	Data Collection for Argument Preparation	Presentation	Decision Making
Man A Course of Study, 1970	Prototype trial of the judicial model applied to curriculum adoption	Judicial	Based on logical analysis by advocates	Analysis of existing documents	Judicial hearing	N.A.
Experience-Based Career Education, 1974	A dissemination strategy for educators considering the adoption of EBCE in their districts	Judicial	Based on interviews and priority ratings by potential adoptors	Prior evaluation reports, program documents and testimony of witnesses	Judicial hearing and videotape	Videotape of hearing was one input to school district administrators considering adoption of EBCE
Independent Plans of Study, 1974	Summative evaluation and input for program continuation decisions	Debate	Issues selected by both teams	Records search, questionnaires, joint interviews	Debate, rebuttal, questions by panel, videotape, meta-evaluation report	Four-person panel made judgments (based solely on the hearing). These served as recommendations to the decision makers.
Division of Teacher Education, 1975	Summative evaluation and input for program continuation decisions	Judicial	Based on interviews and priority ratings of I.U. faculty and students	Prior evaluation reports, program documents, testimony of witnesses	Judicial hearing, panel deliberations, videotape, meta-evaluation report	Jury made judgments (based solely on the hearing) and recommendations to the decision makers
Graduate Program in Clinical-Community Psychology, 1976	Input for program continuation decisions	Judicial	Formal complaint and response issued	Program documents, testimony of witnesses	Judicial hearing, panel deliberations, videotape, meta-evaluation report	Jury made judgments (based solely on the hearing) on four issues and recommendations to the decision makers
Hawaii 3 on 2, 1977	Summative evaluation and input for program continuation decisions	Debate	Based on program objectives and interviews with hundreds of educators, parents, community members, and legislators	Program documents, prior evaluation reports, student cognitive and affective data, interviews, questionnaires, public meetings, classroom observations, cost analysis	Debate, questions by audience, television presentation, statewide public reactions collected, written technical and adversary reports	The debate was one input to the decision makers and public in Hawaii
Twin City Institute for Talented Youth, 1971	Formative evaluation	Position paper	N.A.	N.A.	A 1- or 2-page statement in the written evaluation report	Not specified

N.A. = Not Applicable

decisions. In nonadversary evaluations, the evaluator often assumes that written objectives can be the sole focus for an evaluation.

In each of the six case studies presented, an effort was made to evaluate the evaluation process itself. With the IPS and DTE evaluation, an outsider was hired specifically to conduct such an evaluation. In the case of the EBCE hearing, an outside anthropologist was employed to observe and record notes on the proceedings. In five of the cases, a videotape of the hearings was produced and used to solicit reactions. This general openness to having others assess one's evaluation activities is encouraging.

Some Unresolved Issues

As with any innovation, there are often more questions raised than there are answers. Listed below are some important questions we feel need to be addressed in the years ahead.

1. Is it realistic or desirable for decision makers to make important judgments about a program based on a single hearing?
2. How can evaluation issues be better formulated for a hearing, and how can more effective standards by which to judge these issues be established?
3. How can "hard" data be more effectively integrated with human testimony in adversary hearings?
4. What special legal and investigative skills should be provided in training programs for evaluators in the social sciences?
5. In addition to judicial hearings, debates and adversary position papers, what additional metaphors may be useful in evaluation? (For example, the Congressional hearing model may be very useful.)

6. Can the win-loss implication of an adversary evaluation be modified to allow a more honest portrayal of a program?

7. Frustrations arise if standards for assessing experimental research are applied to adversary evaluation. What, then, are appropriate standards to use in assessing an adversary evaluation?

Because adversary evaluation provides a highly visible format, it is easy to jump on the bandwagon or to reject this approach outright. We hope that you will take a careful look at its potential, weigh its strengths and weaknesses, and then decide if this approach could be useful in your situation.

REFERENCES

House, E.R. Justice in Evaluation. Unpublished manuscript, University of Illinois, Center for Instructional Research and Curriculum Evaluation, 1976.

Levine, M. Scientific Method and the Adversary Model: Some Preliminary Suggestions. American Psychologist, 1974, 31, 661-677.

Levine, M. Experiences in Adapting the Jury Trial to the Problem of Educational Program Evaluation. Unpublished manuscript, State University of New York at Buffalo, 1976.

Northwest Regional Educational Laboratory. 3 on 2 Evaluation Report: Executive Summary. Portland, Oregon: Author, 1977.

Northwest Regional Educational Laboratory. 3 on 2 Evaluation Report: Technical Report. Portland, Oregon: Author, 1977.

Owens, T.R. Educational Evaluation by Adversary Proceeding. In E.R. House (ed.), School Evaluation: The Politics and Process. Berkeley: McCutchan Publishing, 1973.

Stake, R.E. and Gjerde, C. An Evaluation of TCITY: The Twin City Institute for Talented Youth. Urbana, Illinois: University of Illinois, Center for Instructional Research and Evaluation, 1971.

Stenzel, N. Evaluation and the Jury Trial Metaphor: Meta-Evaluation of an Adversary Hearing Evaluation of the Division of Teacher Education at Indiana University. Unpublished manuscript, 1975. (Available from Illinois Office of Education, Department for Exceptional Children, Springfield, Illinois.)

Stenzel, N. Interdisciplinary Perspectives on Adversary Evaluation. Unpublished manuscript, 1976. (Available from Illinois Office of Education, Department for Exceptional Children, Springfield, Illinois.) (a)

Stenzel, N. Meta-Evaluation of the Individual Plans of Study (IPS) Adversary Hearing Evaluation. Unpublished manuscript, 1976. (Available from Illinois Office of Education, Department for Exceptional Children, Springfield, Illinois.) (b)

Wolf, R.L. Trial by Jury: A New Evaluation Method. Phi Delta Kappan, 1975, 57, 185-187.

BIBLIOGRAPHY

Arnstein, G.E. Trial by Jury: A New Evaluation Method. Phi Delta Kappan, 1975, 57, 188-190.

Franklin, M.A. The Biography of a Legal Dispute: An Introduction to American Civil Procedure. Mineola, New York: Foundation Press, 1958.

Garvey, C. and Levine, M. Program Evaluation: On Trial. North Central Association Quarterly, 1976, 50, 359-363.

Hiscox, M.D. and Owens, T.R. Implementing a Prototype Educational Adversary Model. Paper presented at the meeting of the Washington Educational Research Association, Seattle, 1975.

House, E.R. Justice in Evaluation. Unpublished manuscript, University of Illinois, Center for Instructional Research and Curriculum Evaluation, 1976.

Kourilsky, M. An Adversary Model for Educational Evaluation. Evaluation Comment, 1973, 4 (2), 3-6.

Kourilsky, M. and Baker, E.L. An Evaluation of Evaluation Procedures: An Experimental Comparison of Interaction, Advocacy and Adversary Education. Paper presented at the meeting of the American Educational Research Association, Chicago, 1975.

Levine, M. Scientific Method and the Adversary Model: Some Preliminary Suggestions. American Psychologist, 1974, 31, 661-677.

Levine, M. Experiences in Adapting the Jury Trial to the Problem of Educational Program Evaluation. Unpublished manuscript, State University of New York at Buffalo, 1976.

Northwest Regional Educational Laboratory. 3 on 2 Evaluation Report (3 vols.). Portland, Oregon: Author, 1977.

Owens, T.R. Educational Evaluation by Adversary Proceeding. In E.R. House (ed.), School Evaluation: The Politics and Process. Berkeley: McCutchan Publishing, 1973.

Reinhard, D.L. Methodology Development for Advocate Team Use for Input Evaluation. Paper presented at the meeting of the American Educational Research Association, New Orleans, 1973.

Stake, R.E. and Gjerde, C. An Evaluation of TCITY: The Twin City Institute for Talented Youth. Urbana, Illinois: University of Illinois, Center for Instructional Research and Evaluation, 1971.

Stenzel, N. Evaluation and the Jury Trial Netaphor: Meta-Evaluation of an Adversary Hearing Evaluation of the Division of Teacher Education at Indiana University. Unpublished manuscript, 1975. (Available from Illinois Office of Education, Department for Exceptional Children, Springfield, Illinois.)

Stenzel, N. Interdisciplinary Perspectives on Adversary Evaluation. Unpublished manuscript, 1976. (Available from Illinois Office of Education, Department for Exceptional Children, Springfield, Illinois.)

Stenzel, N. Meta-Evaluation of the Individual Plans of Study (IPS) Adversary Hearing Evaluation. Unpublished manuscript, 1976. (Available from Illinois Office of Education, Department for Exceptional Children, Springfield, Illinois.)

Wolf, R.L. The Application of Select Legal Concepts to Educational Evaluation. Unpublished doctoral dissertation. University of Illinois, 1973.

Wolf, R.L. The Citizen as Jurist: A New Mode of Educational Evaluation. Citizen Action in Education, Winter, 1974, p. 4.

Wolf, R.L. Evidence: Educational Evaluation and the Metaphors of Law. Paper presented at the meeting of the American Educational Research Association, Washington, D.C., 1975.

Wolf, R.L. Trial by Jury: A New Evaluation Method. Phi Delta Kappan, 1975, 57, 185-187.